W. Kropatsch, R. Klette, F. Solina (eds.)
in cooperation with R. Albrecht

Theoretical Foundations of Computer Vision

SpringerWienNewYork

W. Kropatsch
R. Klette
F. Solina (eds.)
in cooperation with
R. Albrecht

Theoretical Foundations of Computer Vision

Computing Supplement 11

SpringerWienNewYork

Prof. Dr. W. Kropatsch
Institute for Pattern Recognition
and Image Processing
Technical University Vienna
Austria

Prof. Dr. F. Solina
Faculty of Electrical Engineering
and Computer Science
University of Ljubljana
Slovenia

Prof. Dr. R. Klette
Computer Science Department
Technical University at Berlin
Germany

Prof. Dr. R. Albrecht
Institute for Informatics
University of Innsbruck
Austria

This work is subject to copyright.
All rights are reserved, whether the whole or part of the material is concerned, specifically those of translation, reprinting, re-use of illustrations, broadcasting, reproduction by photocopying machines or similar means, and storage in data banks.

©1996 Springer-Verlag/Wien
Printed in Austria

Typesetting: Asco Trade Typesetters Ltd., Hong Kong
Printing: Fa. Novographic, A-1238 Wien
Binding: Ferdinand Berger & Söhne Ges.m.b.H., A-3580 Horn

Printed on acid-free and chlorine-free bleached paper

With 87 Figures

Library of Congress Cataloging-in-Publication Data

Theoretical foundations of computer vision / W. Kropatsch, R. Klette,
F. Solina, eds. in cooperation with R. Albrecht.
p. cm. – (Computing supplement, ISSN 0344-8029 ; 11)
Includes bibliographical references.
ISBN 3-211-82730-7 (alk. paper)
1. Computer vision. 2. Image processing – Digital techniques.
I. Kropatsch, W. (Walter) II. Klette, Reinhard. III. Solina,
Franc. IV. Series: Computing (Springer-Verlag). Supplementum ; 11.
TA1634.T43 1996
006.4'2 – dc20 95-45033

ISSN 0344-8029
ISBN 3-211-82730-7 Springer-Verlag Wien New York

Contents

Computing Suppl. 11, 201–220 (1996)

Computing

© Springer-Verlag 1996
Printed in Austria

Segmentation with Volumetric Part Models*

F. Solina, Ljubljana

Abstract

Segmentation with Volumetric Part Models. Volumetric models are top-level shape representation in computer vision applications. Volumetric models are especially suited for part-level representation on which manipulation, recognition and other reasoning can be based. The two most popular types of volumetric models in computer vision are generalized cylinders and superquadrics. This paper gives an overview of recovery and segmentation methods applying these two types of volumetric models. Methods of segmentation into parts are analyzed and advantageous properties of part-models discussed.

Key words: Shape representation, superquadrics, generalized cylinders, part recovery.

1. Introduction

The goal of computer vision is to enable intelligent interaction of artificial agents with their surroundings. The means of this interaction are images of various kinds; intensity images, pairs of stereo images, range images or even sonar data. Images which at the sensory level consist of several hundreds or thousands of individual image elements must in this process be encoded in a more compact fashion. For any reasoning or acting on the surroundings, it is advantageous that this coding of images as well as the internal representation of the observed scene closely reflects the actual structure. Distinct objects, for example, should have distinct models of themselves. In this way, the labeling of individual entities, necessary for control and higher level reasoning, becomes possible.

So far, many different models have been used for modeling different aspects of objects and scenes. Models for representing 3D structures can be grouped into local and global models. Methods for local representation attempt to represent objects as sets of primitives such as surface patches or edges. Global methods on the other hand attempt to represent an object as an entity in its own coordinate system. When objects of such global models correspond to perceptual equivalents

* This research was supported in part by The Ministry for Science and Technology of The Republic of Slovenia (Projects P2-1122 and J2-6187).

of parts, we speak of part-level models. Several part-level models are required to represent an articulated object. A part-level shape description supports spatial reasoning, object manipulation, and structural object recognition. People often resort to such part description when asked to describe natural or man-made objects [37]. Such part descriptions are generally suitable for path planning or manipulation—for object-recognition, however, they are sometimes not malleable enough to represent all necessary details and several researchers are looking into extending part-level models with additional layers of details.

To obtain part-level descriptions of a scene *two* tasks must be accomplished. The image must be partioned into areas corresponding to individual parts—a problem referred to as segmentation—and recovering a part model for each of those segments. Normally, these two tasks are separated so that segmentation is performed first and then followed by modeling of previously isolated segments. In this way, segmentation cannot take directly into account the shapes that part-models can adopt. Adequate part-models for all such image segments may not exist in the selected modeling language. To avoid this problem segmentation and part-model recovery can be combined so that images are segmented only into parts which are instantiations of selected part-models.

The two most popular types of volumetric models employed in computer vision are generalized cylinders and superquadrics. Several other global models exist, however, that attempt to represent an object as an entity in their own coordinate system: Spherical harmonic surfaces [43], Gaussian images and extended Gaussian images [26], Symmetry seeking models [49], Blobby model [35], and Hyperquadrics [23]. In this paper we concentrate on the two most commonly used models: generalized cylinders and superquadrics. We compare them with regard to recovery and segmentation. We show that superquadrics are in respect to recovery from images advantageous in comparison to generalized cylinders, especially when segmentation and model recovery are to be combined.

The rest of the paper is divided as follows: the second section is on segmentation and ways of combining segmentation with model recovery. The following two sections describe model recovery and segmentation techniques for generalized cylinders and superquadrics. The fifth section discusses some typical applications of volumetric models in computer vision. In the last section general properties of part-level models are discussed.

2. Segmentation

Basic scientific methodology instigates decomposition of complex objects into parts, units or primitives to enable its study at various levels of abstraction. Abstraction is a crucial mechanism to cope with limits on how much information one can process at a time. In a similar way, to comprehend images, they should be decomposed into "natural" and "simple" parts that order and partition visual information into a limited number of perceptually significant parts. This challenging problem is called segmentation. Essential to segmentation is that the resulting parts ought to correspond to the underlying physical part structure of the scene depicted in the image. This is a prerequisite for image understanding. Part-level description of an image is therefore a necessary step towards building the scene description in terms of symbolic entities.

Segmentation entails decomposing images into segments so that each piece of information in an image is mapped either to a segment or discarded as noise. To get as compact a description as possible a minimum number of such part primitives should be used. To define what is "natural" and "simple" is a hard problem. It depends on the type of the observed scene as well as on the objective of the observing agent. The general segmentation problem is very difficult since multiple sources of image information should be involved. In this article we restrict the problem to shape information alone. In absence of domain knowledge, ambiguities can arise due to multiple representations, incomplete data, and multiple degrees of freedom of part-models.

In general, a criterion for segmentation must be defined. Standard segmentation criteria are different measures of homogeneity or difference since the two basic approaches to segmentation are:

1. finding of homogeneous regions, and
2. finding of borders (differences) between regions.

Segmentation methods based on homogeneity criterion can be implemented using different techniques, from simple tresholding, region growing, split and merge to scale-space approaches. Border-based segmentation employs different edge detection methods such as gradient methods, Hough transform, and active contour models.

To reach the part-model level representation that we discussed in the introduction, tokens that were obtained with such segmentation methods must be evaluated, filtered, grouped, and combined. Perfect segmentation into part-models would result only in segments that can be described with selected part-models, be it generalized cylinders or superquadrics. A common problem, however, is that the above-mentioned low-level segmentation methods often give part boundaries that cannot be modeled adequately with selected part-models. The cause of this problem is that grouping and combining of the low-level image elements do not evaluate the results on the basis of the final part-level model but on some lower level model or criterion (lines, contours, cross-sections, symmetries, aspect ratios, compactness etc.).

To overcome this problem, Bajcsy et al. [3] argued that segmentation and part-level modeling should be combined. A general method for combining segmentation and shape-recovery, called "recover-and-select" was developed by Aleš Leonardis [28]. In this article we show that when segmentation is combined with part-model recovery, low-level segmentation can be skipped over and fitness to given part-level shape models used directly as a segmentation criterion.

Hence, we divide segmentation methods into two general categories:

1. *Segment-then-Fit* methods, and
2. *Segment-and-Fit* methods.

Most standard segmentation methods belong to the first category which separates segmentation and shape recovery. This separation accounts for the above discussed problem. The methods in the second category use the final shape models also as a segmentation criterion and in this way achieve better segmentation results.

As mentioned in the Introduction we concentrate in this article on two types of part-level models: generalized cylinders and superquadrics. In the following two sections, which describe segmentation and part-modeling approaches using generalized cylinders and superquadrics, we will try to apply the above defined classification of segmentation methods.

3. Generalized Cylinders

The first dedicated volumetric models in computer vision were generalized cylinders. Generalized cylinders, sometimes referred to also as generalized cones, were defined by Thomas O. Binford in 1971 [8]. Generalized cylinder representation was preceded by earlier concepts for volumetric representation, notably the symmetric axis representation, developed by Blum [9]. Symmetric axis descriptions were defined as the set of centers of maximal spheres contained within the shape. All these representations are especially good at representing naturally evolved or grown elongated shapes.

A generalized cylinder is expressed by a volume obtained by sweeping a two-dimensional set or volume along an arbitrary space curve (Fig. 1). The set may vary parametrically along the curve. Different parameterizations of the above definition are possible. In general, a definition of the axis and the sweeping set are required. The axis can be represented as a function of arc length s in a fixed coordinate

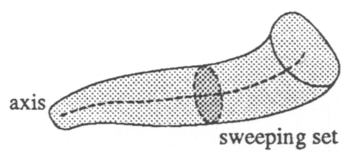

Figure 1. A generalized cylinder is constructed by sweeping a closed contour along a space curve

system x, y, z

$$a(s) = (x(s), y(s), z(s)).$$

The sweeping set is more conveniently defined in a local coordinate system, defined at the origin of each point of the axis $a(s)$. The sweeping set can be defined by a cross section boundary, parameterized by another parameter r:

$$\text{sweeping set} = (x(r,s), y(r,s)).$$

This general definition is very powerful and a large variety of shapes can be described with it. In the most general form the generalized cylinder representation is so powerful that almost arbitrarily formed complete objects can be modeled. But since their outset generalized cylinders were used mainly as part-models [33], especially if restrictions to the general definition were applied (Fig. 2). To limit the complexity and simplify the recovery of generalized cylinder models from images researchers often restricted generalized cylinders to straight axes with constant sweeping sets. Properties of straight homogeneous general cylinders are addressed in [41].

Various methods for deriving descriptions based on generalized cylinders mostly from range images were attempted. One of the most widely publicized vision systems based on generalized cylinders was the ACRONYM system by Brooks [13]. The ACRONYM system was capable of recognizing different types of airplanes from intensity images of airports taken from the air.

Generalized cylinders influenced much of the model-based vision research in the past two decades. Besides building actual vision systems, generalized cylinders had some impact also on vision theory. GEONS introduced by Biederman [7] were conceived as a limited set of different primitive building blocks that could build any natural or man-made shape. The concept of geons was derived from qualitative changes of generalized cylinders. Geons are classified only on the basis of axis shape, cross-section shape, cross-section sweeping function, and cross-section symmetry. These qualitative geometrical properties could prove to be very useful in indexing object databases.

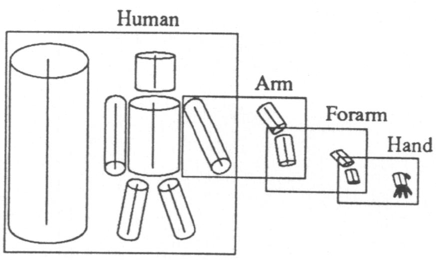

Figure 2. A hierarchical part-level representation of a human form with generalized cylinders as part-models (after [36])

3.1. Recovery of Generalized Cylinders

Recovery of generalized cylinders was undertaken from range images, pairs of stereo images, and from contours derived from single intensity images. Especially recovery from contours received the widest interest because humans are so good at perceiving a shape from its boundary.

Recovery of generalized cylinders has been approached from two broad perspectives:

1. In the first approach, some property of the 3-D surface is associated with each interpretation. The interpretation associated with the highest value of the selected property is chosen. Different properties were proposed as the preference criterion: *smoothness of the curve* [6], *curvature* [53], and *compactness* [12]. The smoothness property is a combination of the curvature and torsion of a 3-D curve. Over all possible 3-D curves that can generate a given 2-D image curve, the smoothest one is selected, or alternatively, the most compact 2-D shape. These methods in general require smooth and complete boundaries, some even implicitly assume planar contours.
2. The second approach to recovery of generalized cylinders are constraint-based. Constraints on 3-D surface orientations are inferred from a variety of observations (skew and parallel symmetry [27, 51]) and their propagation to other parts of the image (tilt and slant [48]). A unique solution is expected when various such constraints are combined.

Most recovery methods are restricted to straight homogeneous general cylinders (SHGC's) [51]. In general, one can criticize the methods of recovering generalized cylinders on the count that they are not very robust because they must rely on complicated rules for grouping of low level image models (i.e. edges, corners, surface normals) into models of larger granularity (i.e. symmetrical contours or cross-sections) to arrive finally to generalized cylinders. The recovery methods are sensitive to noise and often require complete data without any occlusions. These problems are due in part to the complicated parameterization of generalized cylinders and to the lack of a fitting function that would enable a straightforward numerical examination of the model's appropriateness for the modeled image data.

3.2. Segmentation with Generalized Cylinders

Most systems that recover generalized cylinders separate segmentation and model recovery and can be classified as "Segment-then-Fit" methods. The ACRONYM system, for example, finds first "ribbons", which are two-dimensional specializations of generalized cylinders, by an edge linking algorithm. By a system of forward and backward constraints these ribbons can be later matched to generalized cylinders which make up the model base.

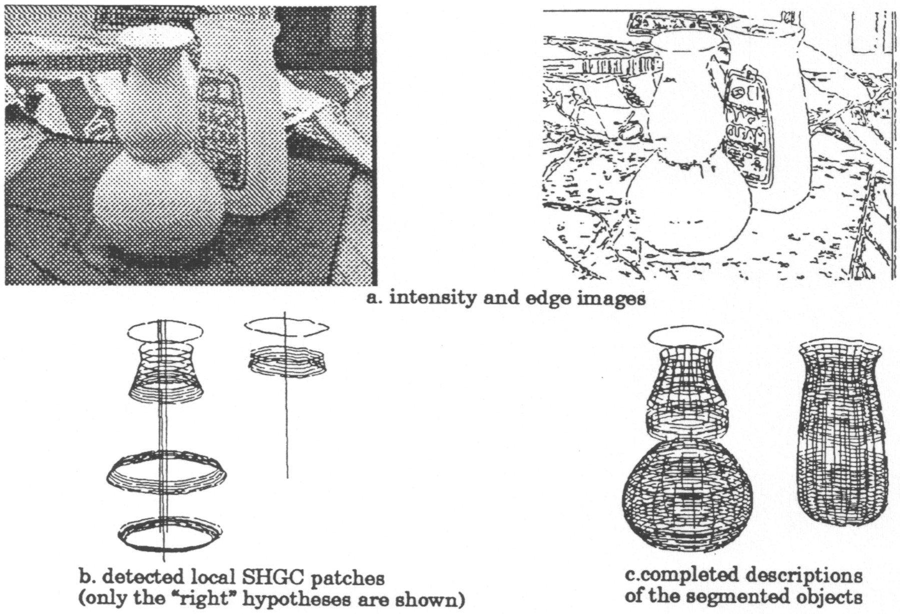

Figure 3. Combined segmentation and recovery of straight homogeneous generalized cylinders (from [56])

Since generalized cylinders do not have a corresponding implicit function, a direct numerical evaluation of the model's fitness to the data is very demanding. This is why an integrated approach to shape recovery and segmentation is difficult. Notwithstanding these obstacles, the latest results in recovery and segmentation of straight homogeneous generalized cylinders (SHGC) achieved by Zerrough and Nevatia are encouraging [56] (Fig. 3). The method which integrates segmentation and recovery of SGHCs is quite complex and requires three grouping levels: the curve level, the parallel symmetry level, and the SHGC patch level. The SHGC level is intended to form complete SHGC object descriptions whenever possible in the image. Constraints used in all three grouping levels are derived from geometric projective properties of SHGCs.

4. Superquadrics

Superquadric models appeared in computer vision as an answer to some of the problems with recovery of generalized cylinders [37]. Superquadrics are solid models that can, with a fairly simple parameterization, represent a large variety of standard geometrical solids as well as smooth shapes in between. This makes them convenient for representing rounded, blob-like shaped parts, typical for objects formed by natural processes.

To introduce the concept of superquadrics we define their 2-D equivalent. A superellipse is a closed curve defined by the following simple equation:

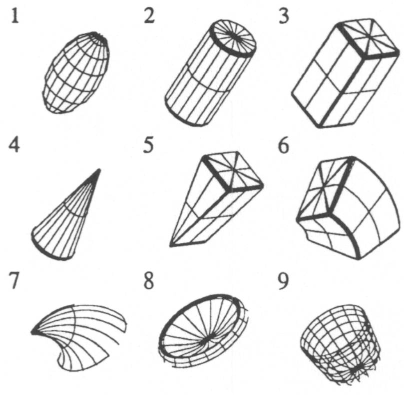

Figure 4. Superquadric models enhanced with global deformations (from [47])

$$\left(\frac{x}{a}\right)^m + \left(\frac{y}{b}\right)^m = 1,$$

where a and b are the size (positive real number) of the major and minor axes and m is a rational number

$$m = \frac{p}{q} > 0, \quad \text{where} \begin{cases} p & \text{is an even integer,} \\ q & \text{is an odd integer.} \end{cases}$$

If $m = 2$ and $a = b$, we get the equation of a circle. For larger m, however, we get gradually more rectangular shapes, until for $m \to \infty$, the curve takes up the shape of a square. Superellipses are special cases of curves which are known in analytical geometry as Lamé curves [31][1]. Piet Hein, who popularized these curves for design purposes, also made a generalization to 3D which he named *superellipsoids* or *superspheres* [17]. The final mathematical foundation of superquadrics was laid out by Barr [4], who generalized the whole family of quadric surfaces with the help of varying exponents, and coined a new name for them—*superquadrics*. Superquadrics are by definition a family of shapes that includes not only superellipsoids, but also superhyperboloids of one and of two pieces, as well as supertoroids (see Fig. 4).

The explicit superellipsoid equation, defined by the following surface vector, is

$$\mathbf{x}(\eta, \omega) = \begin{bmatrix} a_1 \cos^{\varepsilon_1}(\eta) \cos^{\varepsilon_2}(\omega) \\ a_2 \cos^{\varepsilon_1}(\eta) \sin^{\varepsilon_2}(\omega) \\ a_3 \sin^{\varepsilon_1}(\eta) \end{bmatrix} \qquad \begin{matrix} -\pi/2 \le \eta \le \pi/2 \\ -\pi \le \omega < \pi \end{matrix}, \tag{1}$$

[1] Lamé curves are named after the French mathematician Gabriel Lamé, who was the first who studied them in *Examen des différentes méthodes employées pour résoudre les problémes de geometrie*, Paris, 1818.

where a_1, a_2 and a_3 determine size, and ε_1 and ε_2 determine global shape. The alternative, implicit superellipsoid definition, also called the *inside-outside* function is

$$\left(\left(\frac{x}{a_1}\right)^{2/\varepsilon_2} + \left(\frac{y}{a_2}\right)^{2/\varepsilon_2}\right)^{\varepsilon_2/\varepsilon_1} + \left(\frac{z}{a_3}\right)^{2/\varepsilon_1} = 1. \tag{2}$$

Points x, y, z that correspond to the above equation are on the surface of the superellipsoid.

For numerical calculation, it is easier to assume that exponents ε_1 and ε_2 can be any positive real number and not only rational numbers with an even enumerator. Then, one should assume that exponentiation in Eqs. (1) and (2) means

$$x^p = \mathrm{sign}(x)|x|^p = \begin{cases} x^p & x \geq 0 \\ -|x|^p & x < 0 \end{cases}$$

to avoid complex numbers when a negative real number is raised to a real exponent. For applications in computer vision, the values for ε_1 and ε_2 are normally bounded: $2 > \varepsilon_1, \varepsilon_2 > 0$, so that only convex shapes are produced. For a superquadric in canonical position one needs to set the value of 5 parameters (3 for size in each dimension, 2 for shape defining exponents). For a superquadric in general position 6 additional parameters are required to define the translation and rotation of the model.

Superquadric models, which compactly represent a continuum of useful forms with rounded edges, and which can easily be rendered and shaded due to their dual normal equations, and deformed by parametric deformations, are very useful in computer graphics. Parametric deformations such as twisting, bending, tapering, and their combinations can enhance the expressive power of superquadrics [5]. Parametric deformations typically require just a few more parameters.

Pentland [37] was the first who grasped the potential of the superquadric models and parametric deformations for modeling natural shapes in the context of computer vision. He proposed to use superquadric models in combination with global deformations as a set of primitives which can be molded like lumps of clay to describe the scene structure at a scale that is similar to our naive perceptual notion of *parts*. Pentland presents several perceptual and cognitive arguments to recover the scene structure at such a part-level since people seem to make heavy use of this part structure in their perceptual interpretation of scenes. The superquadrics, which are like phonemes in this description language, are deformed by stretching, bending, tapering or twisting, and then combined, using Boolean operations to build complex objects. In general, the same arguments as the ones for using generalized cylinders as part models hold for superquadrics.

Superquadrics are in fact a subset of generalized cylinders. Any superquadric can be in principle represented as a generalized cylinder admitting that the parameterization could be much more complicated. The geon concept can also be expressed in terms of superquadrics. Raja and Jain [42] conducted experiments on mapping superquadric shapes to 12 shape classes corresponding to a "collapsed" set of 36 different geons. Raja and Jain used the five shape and deformation parameters of superquadrics for classification into 12 geon classes using binary tree and k-nearest-neighbor classifiers.

4.1. Recovery of Superquadrics

The problem of recovering superquadrics from images is an overconstrained problem. A few model parameters (i.e. 11 for non-deformed superquadrics) must be determined from several (i.e. a few hundred) image features (range points, surface normals or points on occluding contours). By its parameterization the superquadrics impose a certain symmetry and in this way place some reasonable constraints on the shape of the occluded portion of a three dimensional object.

In the first article on the use of superquadrics in computer vision, Pentland [37] proposed an analytical method for recovery of superquadrics using the explicit equation (1). Except for some simple synthetic images, this analytical approach did not turn out to be feasible. Pentland [38] later proposed another method which combined recovery with segmentation and was based on a coarse search through the entire superquadric parameter space for a large number of overlapping image regions. The major objection to this method is its excessive computational cost.

Iterative methods based on non-linear least squares fitting techniques using different distance metrics were proposed [2, 10]. Solina and Bajcsy [2, 47] formulated the recovery of deformed superquadric models from pre-segmented range data as a least-squares minimization of a fitting function. An iterative gradient descent method was used to solve the non-linear minimization problem. A modified superquadric implicit or inside-outside function (Eq. (2)) with an additional multiplicative volume factor was used as the fitting function. The volume factor is used to ensure the recovery of the *smallest* superquadric model that fits the range data in the least squares sense. To the standard superquadric model, which requires 11 parameters, linear tapering, bending, and a cavity deformation were added, which adds up to a total of 18 parameters. Recovery of a single superquadric model requires on the average about 30 iterations (see Fig. 5).

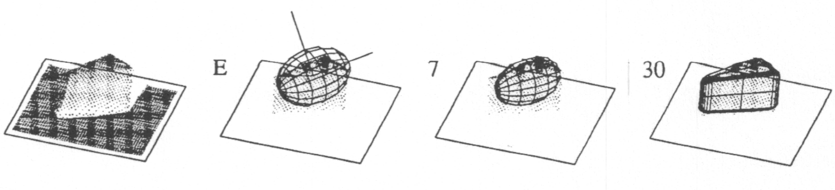

Figure 5. Recovery of a tapered superquadric from pre-segmented range data (from [47]). From left to right: original range image; *E* initial estimate; *7, 30* models after the 7th and 30th iteration

Pentland [39] proposed another superquadric recovery method. Segmentation was first achieved by matching 2D silhouettes (2D projections of 3D superquadric parts of different shapes and of different orientations) to the image data. After part segmentation, superquadric models were fitted to range data of individual part regions. Superquadric fitting based on modal dynamics [40] used as the error metric the squared distance along the depth axis z between the range data and the projected volume's visible surface.

Hager [22] combined the estimation (recovery) process of superquadric models with the decision-making process (i.e. graspability, categorization). Usually both processes are divorced in the sense that first a recovery process is performed and then a decision is made based on the recovered models. Combining both stages should result in minimal work required to reach a decision. The approach is based on a interval-bisection method to incorporate sensor based decision making in the presence of parametric constraints. The constraints describe a model for sensor data (i.e. superquadric) and the criteria for correct decisions about the data (i.e. categorization—see also [46]). An incremental constraint solving technique performs the minimal model recovery required to reach a decision. The major drawback of the method is slow convergence when categorization is involved. Determining the shape parameters ε_1 and ε_2 required several hundred iterations.

Yokoya et al. [57] experimented with simulated annealing to minimize a new error-of-fit measure for recovery of superquadrics from pre-segmented range data. The measure is a linear combination of distance of range points to the superquadric surface and difference in surface normals (first proposed by Bajcsy and Solina in [2]). Several hundred iterations were needed to recover models from range data.

Vidmar and Solina [52] studied the recovery of superquadrics from 2D contours. For a given contour several possible superquadric interpretations were derived. To a human observer some of these interpretations are obviously more natural than others, although all recovered models have a very tight fit to the contour data. Perceptually better solutions could be selected by using just a few additional pieces of information (a few range points or shading information).

Horikoshi and Suzuki [25] multiplied the objective function with a weighting function for robust estimation (based on whether the point is closer to the median value of the inside-outside function, or far from it in either directions). Consequently, the model is less sensitive to outliers.

Since no direct comparison of different metrics and minimization methods for superquadric recovery was made, it is difficult to rank the recovery methods only on the basis of results presented in the articles. Some experimental comparisons of different error-of-fit measures are given in [18]. Gupta [20] also discusses the error-of-fit functions. What is finally important is the perceptual likeness of models to the actual objects, the speed of convergence, and last but not least, the simplicity of implementation. On this ground the method proposed in [47] received a wide acceptance since several other authors have used it in their vision or robotic systems [1, 16, 19, 20, 30, 42].

4.2. *Segmentation with Superquadrics*

The early work on superquadrics concentrated primarily on recovery of isolated parts and did not address segmentation or used superquadrics as volumetric primitives *after* segmentation was achieved. Once recovery of superquadrics was well understood, more sophisticated techniques were designed that applied superquadrics to scene segmentation [19, 25, 30].

4.2.1. Two-Stage Segment-then-Fit Methods

These two stage methods decouple segmentation and model recovery. First, images are segmented, then part-models are fitted to the resulting regions.

Pentland [39], for example, used matched filters to segment binarized image data into part regions. The best set of binary patterns that would completely describe a silhouette is selected. The 3D data corresponding to each of the selected patterns was then fitted with a deformable superquadric based on modal dynamics.

Gupta et al. [21] used an edge based region growing method to segment range images of compact objects in a pile. The regions were segmented at jump boundaries, and each recovered region was considered a superquadric object. Reasoning was done about the physical support of these regions, and several possible 3D interpretations were made based on various scenarios of the object's physical support. A superquadric model was fitted and classified corresponding to each recovered object.

Ferrie et al. [16] used differential geometric properties and projected space curves modeled as snakes for segmenting range data. An augmented Darboux frame is computed at each point by fitting a parabolic quadric surface, which is iteratively refined by a curvature consistency algorithm.

Another qualitative shape recovery method using geon theory was proposed by Metaxas and Dickinson [34] to recover superquadrics on intensity data. Their integrated method uses Dickinson et al.'s geon-based segmentation scheme [14] (into ten geon classes) to provide orientation constraint and edge segments of a part. This is then the input to the physically-based global superquadric model recovery scheme developed by Terzopoulos and Metaxas [50].

All two-stage approaches suffer from the problem that the results of segmentation might not correspond tightly to any superquadric model. Thus, model recovery on such part domain will be uncertain about the shape, size and orientation of the model. To describe adequately a scene with a particular shape language, one should use this language also for partitioning the scene. In this case, a method that can accommodate the segments to the orientation, shape, and size of the superquadric model must be used. This can be accomplished by interleaving model recovery and segmentation [3].

4.2.2. Interleaved Segment-and-Fit Methods

Solina first attempted part-segmentation with superquadrics by recursively splitting the modeling domains and rejecting extraneous range points during model recovery [44]. It was, however, extremely difficult to constrain the single model recovery to take part-structure into account. Clearly, segmentation into part models must recover parts by hypothesizing parts and testing (evaluating) them.

Pentland [38] was the first one to successfully integrate segmentation and superquadric model recovery. However, the brute-force method of searching the entire parameter space of superquadrics for a large number of overlapping image regions is computationally excessively expensive.

Gupta and Bajcsy [19], proposed a recursive global-to-local technique for part recovery. A global model is recovered for the entire data (using [47]), which is evaluated by studying local and global residuals so that the further course of action can be determined. A set of qualitative acceptance criteria define suitability of the model. If the model is found to be deficient in representing data, then additional part models are hypothesized on the un-described data (regions of surface underestimation). The global model is refitted on the remaining data. Thus, the global model shrinks while the part models grow, yielding a hierarchical part-structure.

A tighter integration of segmentation and model recovery was achieved by combining the "recover-and-select" paradigm developed by Leonardis [28, 29] with the superquadric recovery method by Solina [44]. This work demonstrates that superquadrics can be *directly* and *simultaneously* recovered from range data without the mediation of any other geometric models [30]. The *recover-and-select* paradigm for the recovery of geometric parametric structures from image data [28] was originally developed for the recovery of parametric surfaces [29]. The paradigm works by recovering independently superquadric part models everywhere on the image, and selecting a subset which gives a compact description of the underlying data (Fig. 6).

Horikoshi and Suzuki [25] proposed a segment-and-merge method to segment 2D contours (with the figure of interest separated from the background) and sparse 3D data. This recursive procedure results in a possibly overlapping convex superquadric parts. Parts are then merged to arrive at a compact description.

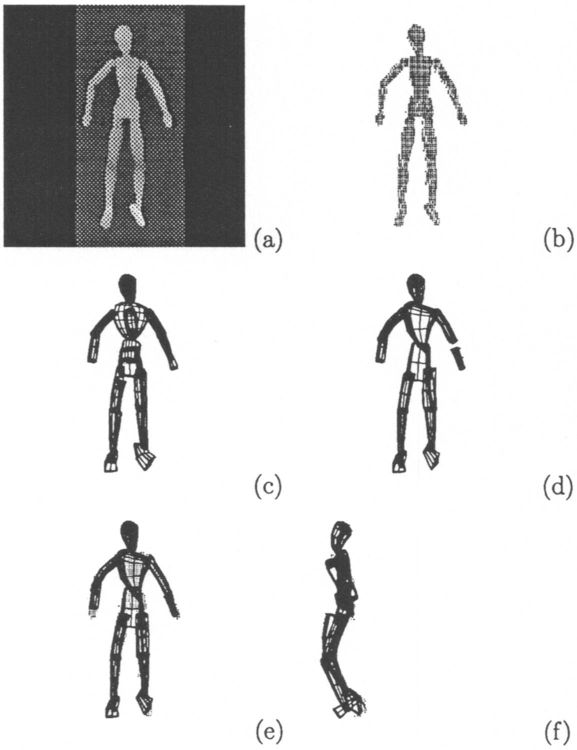

Figure 6. Range image segmentation using deformed superquadric part models. The human form was segmented by employing the *recover-and-select* paradigm of Leonardis [28] and the superquadric recovery method of Solina [44]. **a** Original range image, **b** input range points, **c**, **d** models during the recovery process, **e**, **f** the final result from two views

5. Application of Volumetric Models in Computer Vision

This section explores the role of volumetric shape primitives in computer vision tasks. Typical tasks that require volumetric shape primitives are object classification and object recognition. For simple object classification tasks which are required, for example, for grasping, sorting or object avoidance, part-models offer sufficient information.

The role of part-models in recognition depends on the type and number of objects that need to be recognized. Sometimes the structure and coarse shape of individual parts are enough. Often, the surface detail that the discussed part-models offer is poor for the required recognition task. As is the case with any shape primitive, superquadrics as well as straight homogeneous generalized cylinders have a limited shape vocabulary. They can be used to capture the global coarse shape of a 3D object or its constituent parts. The addition of global deformations increases the expressive power of superquadrics, but still limits it to the global coarse shape as opposed to local details. This lack of fine scale representation can be addressed by adding local degrees of freedom to superquadrics [40, 50], or relax the usual constraints on generalized cylinders. Terzopoulos et al. proposed a "symmetry-seeking model" for 3D object reconstruction which is a kind of a generalized cylinder that can accommodate even local "bumps" on the overall shape [49].

However, one drawback of such locally deformable extensions is that they have too many degrees of freedom to meaningfully segment even a simple scene. The increase in expressive power of part-models also results in an increase in complexity of all the visual tasks like segmentation, representation, recognition, and classification. Consequently, all of the segmentation and classification work with superquadrics [16, 19, 30, 39] has used globally deformable models, limiting the role of local deformations to refine the surface details.

The earliest works on superquadrics dealt primarily with single model analysis, since interpretation of a complex scene required model recovery to be understood first. These methods focussed either on classification of single models, where the power of superquadrics as a compact parametric model was exploited [24, 42, 46], or on using superquadrics as a volumetric primitive *after* a segmentation had been obtained [15, 21, 38]. Once the model recovery was understood, more sophisticated techniques were designed to apply superquadrics to scene segmentation [19, 25, 30].

Solina and Bajcsy [46] recovered objects in the postal domain and categorized them as flats, tubes, parcels, and irregular packages based on the shape and size parameters of the segmented recovered models. Gupta et al. [21] extended Solina's approach to work on a cluttered scene by segmenting the range image using an independent edge-based scheme, and then recovering individual postal objects after reasoning about the physical supporting plane to constrain the 3D shape of the object.

Horikoshi and Kasahara [24] partitioned the superquadric parametric space between $0 < \varepsilon_1 \leq 2.0$ and $1.0 \leq \varepsilon_2 \leq 3.0$ to develop a shape indexing language. They mapped the representation space to verbal instructions like "rounder", "pinch", "flatten", etc., and developed a man-machine interface to construct object models. They also described an indexing scheme where complex objects were stored as superquadric models and indexed by model parameters.

Model-driven recognition (with superquadric part-primitives) has not so far been exploited despite the compact representation. The reason is that it is very difficult to recover "canonical" representations of objects from real data. Instead of recovering canonical descriptions, most researchers have followed the data-driven bottom-up strategy of fitting superquadric models to the data. The recognition problem then reduces to matching the recovered superquadric parts with the superquadric parts in the model base.

6. Conclusions

Part-modeling is a convenient abstraction mechanism in image understanding and practiced in many different kinds of applications. This article focused in particular on two types of part-models: generalized cylinders and superquadrics. Methods for their recovery and their role in segmentation were reviewed.

Recovery of generalized cylinders and segmentation with generalized cylinders is hampered by complicated methods that require grouping and combination of various constraints which results in relatively slow implementations and this even with confined forms of generalized cylinders (SHGCs). Most of the research on recovery of generalized cylinders used contours as input.

Despite initial reluctance in using superquadrics due to their nonlinear form, they have proven to be the primitives of choice for many a researcher seeking a volumetric model. The interchangeable implicit and explicit superquadric equations enable numerical evaluation of the model directly on the range data and computation of other desirable shape properties (i.e. surface normals). Superquadric models have shown to be useful as volumetric shape primitives for object categorization, segmentation, recognition, and representation. "Segment-and-fit" method using superquadrics reports good results on objects which would otherwise be unsegmentable with surface-based techniques [30]. Interpretation of intensity and sparse 3D data with superquadrics is still an open problem.

For a fair comparison of generalized cylinders and superquadrics, straight homogeneous generalized cylinders and superquadrics enhanced with global deformations should be used since for these two types of part-models methods of recovery exist. To evaluate and compare different representations the following three criteria defined by Marr and Nishihara [33] can help:

1. *accessibility—can the desired description be computed from an image and can it be done economically*? Faster, conceptually more concise, and more reliable methods exist for recovery of superquadrics.
2. *scope and uniqueness—what class of images is the representation designed for*? Due to their capability of modeling rounded edges and corners superquadrics enhanced with global deformations offer a richer vocabulary for description of naturally occurring shapes than do SHGCs. On the other hand are superquadrics, unlike generalized cylinders, limited in the shape of the cross section along the main axis.
3. *stability and sensitivity—do differences between descriptions in the representation reflect the relative importance of differences between the images described with respect to the task at hand*? Generalized cylinders and superquadrics offer a relatively stable part-level decomposition. Although the shape of recovered part-models is normally stable, the local coordinate system is not necessarily stable. This means that the actual values of model parameters cannot be used directly for recognition. Sensitivity depends on the required level of detail and is coupled with the scale of the representation.

Using a parametric shape model for vision requires that model evaluation be built into the segmentation and recognition systems. To this end, it is imperative to study the residuals of shape models by comparing them against the given data. Whaite and Ferrie [55] have recently described a decision theoretic framework to evaluate the fitted models. They extended their earlier work [54] on uncertainty in model parameters to develop three lack-of-fit statistics. Gupta and Bajcsy [19] used global *and* local distribution of residuals to determine model fitness and segmentation options on static data.

Brady [11] proposed three additional criteria for evaluating representations:

1. *rich local support—representation should be information preserving and locally computable* (*local frames*). Generalized cylinders satisfy this since there is a natural local coordinate frame at all points along the spine [11]. Superquadrics are computable even on small patches of range data, called seeds in [30].
2. *frames—by smooth extension and subsumption local frames should give rise to more global descriptions called frames.* Smooth extension and subsumption is implicit in the definition of generalized cylinders [11]. How can this be exploited for recovery of generalized cylinders is unclear. A superquadric can extend (grow) in the process of segmentation until it reaches in the data the natural part limits which are allowed by its internal parameterization. Subsumption can be achieved by a selection mechanism based on the Minimal Description Criteria [30].
3. *propagation of frames—frames that correspond to perceptual sub parts of a shape can be propagated by inheritance or affixment.* In general, there are partially defined frames naturally associated with the general cylinders or superquadrics. The frame is partial in that certain choices are underconstrained or arbitrary. By inheritance or affixment constraints from adjoining parts can be propagated. In the segmentation method presented in [30] this is not implemented yet.

Real situations demand that the parametric shape models must be recovered on partial and noisy single viewpoint data. Data could be missing due to other objects occluding the view, or due to the shadows in scanner geometry, or due to self-occlusion in single viewpoint data [32]. Noise in 3D measurements is inevitable and most difficult to model. While the symmetry constraints of superquadrics are useful in predicting the missing information, the downside of parametric models is the lack of uniqueness in describing incomplete and noisy data within an acceptable error of tolerance. This fact is borne out in the experiments conducted by [54], where they derive an ellipsoid of confidence within which all the acceptable models lie. Therefore we propose two additional criteria to judge future shape representations:

1. *propagation of uncertainties*—the shape representation should support the representation and propagation of sensor and modeling uncertainties.
2. *use of perceptual preferences in case of insufficient data*—when due to occlusion or sensor characteristics the image data does not offer enough constraints, perceptual preferences should be utilized to restrain the model.

To summarize, because superquadrics have a nice and complete mathematical definition, more powerful mathematical tools are available for their recovery from images. Study of generalized cylinders is distinguished by theoretical rigor but lacks for now efficient methods for recovery from images.

References

[1] Allen, P. K., Michelman, P.: Acquisition and interpretation of 3-D sensor data from touch. IEEE Trans. Robotics Automation *6*, 397–404 (1990).

[2] Bajcsy, R., Solina, F.: Three dimensional shape representation revisited. In: Proceedings of the First International Conference on Computer Vision, pp. 231–240. London: IEEE 1987.

[3] Bajcsy, R., Solina, F., Gupta, A.: Segmentation versus object representation—are they separable? In: Analysis and interpretation of range images (Jain, R., Jain, A., eds.), pp. 207–223. New York: Springer 1990.

[4] Barr, A. H.: Superquadrics and angle-preserving transformations. IEEE Comput. Graphics Appl. *1*, 11–23 (1981).

[5] Barr, A. H.: Global and local deformations of solid primitives. Comput. Graphics *18*, 21–30 (1984).

[6] Barrow, H. G., Tenenbaum, J. M.: Interpreting line drawings as three dimensional surfaces. Art. Intell. *17*, 75–116 (1981).

[7] Biederman, I.: Human image understanding: recent research and theory. Comput. Vision Graphics Image Proc. *32*, 29–73 (1985).

[8] Binford, T. O.: Visual perception by computer. In: Proceedings of the IEEE Conference on Systems and Control, Miami, FL, 1971.

[9] Blum, H.: Biological shape and visual science (Part I). J. Theor. Biol. *38*, 205–287 (1973).

[10] Boult, T. E., Gross, A. D.: Recovery of superquadrics from depth information. In: Proceedings of Spatial Reasoning and Multi-Sensor fusion workshop, pp. 128–137. St. Charles: SPIE 1987.

[11] Brady, M.: Criteria for representations of shape. In: Human and Machine Vision (Beck, J., Hope, B., Rosenfeld, A., eds.), pp. 39–84. New York: Academic Press 1983.

[12] Brady, M., Yuille, A.: An extremum principle for shape from contour. IEEE Trans. Pattern Anal. Machine Intell. *6*, 288–301 (1984).

[13] Brooks, R.: Model-based 3-D interpretation of 2-D images. IEEE Trans. Pattern Anal. Machine Intell. *5*, 140–150 (1983).

[14] Dickinson, S. J., Pentland, A. P., Rosenfeld, A.: From volumes to views: An approach to 3-D object recognition. CVGIP: Image Understanding *55*, 130–1154 (1992).

[15] Ferrie, F. P., Lagarde, J., Whaite, P.: Darboux frames, snakes, and superquadrics: Geometry from the bottom-up. In: Proc. IEEE Workshop on Interpretation of 3D Scenes, pp. 170–176, Austin, TX, 1989.

[16] Ferrie, F. P., Lagarde, J., Whaite, P.: Darboux frames, snakes, and superquadrics: Geometry from the bottom up. IEEE Trans. Pattern Anal. Machine Intell. *15*, 771–784 (1993).

[17] Gardner, M.: The superellipse: a curve that lies between the ellipse and the rectangle. Sci. Amer. *213*, 222–234 (1965).

[18] Gross, A. D., Boult, T. E.: Error of fit measures for recovering parametric solids. In: Proceedings of the 2nd International Conference on Computer Vision, pp. 690–694. Tampa: IEEE 1988.

[19] Gupta, A., Bajcsy, R.: Volumetric segmentation of range images of 3-D objects using superquadrics. CVGIP: Image Understanding, *58*, 302–326 (1993).

[20] Gupta, A., Bogoni, L., Bajcsy, R.: Quantitative and qualitative measures for the evaluation of the superquadric models. In: Proc. IEEE Workshop on Interpretation of 3D Scenes, pp. 162–169. Austin, TX, 1989.

[21] Gupta, A., Funka-Lea, G., Wohn, K.: Segmentation, modeling and classification of the compact objects in a pile. In: Proceedings of the Conference on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques, pp. 98–108, Philadelphia, PA, November 1989. St. Charles: SPIE.

[22] Hager, G. D.: Constraint solving methods and sensor-based decision making. In: Proceedings of the International Conference on Robotics and Automation, pp. 1662–1667, Nice, France, 1992. London: IEEE.

[23] Hanson, A. J.: Hyperquadrics: smoothly deformable shapes with complex polyhedral bounds. Comput. Vision Graphics Image Proc. *44*, 191–210 (1988).

[24] Horikoshi, T., Kasahara, H.: 3-D shape indexing language. In: Proceedings of the International Conference on Computers and Communications, Scottsdale, AZ, 1990. London: IEEE.
[25] Horikoshi, T., Suzuki, S.: 3D parts decomposition from sparse range data using information criterion. In: Proceedings of the CVPR, pp. 168–173, 1993.
[26] Horn, B. K. P.: Robot vision. Cambridge/Mass.: MIT Press 1986.
[27] Kanade, T.: Recovery of the three-dimensional shape of an object from a single view. Art. Intell. *17*, 409–460 (1981).
[28] Leonardis, A.: Image analysis using parametric models: model-recovery and model-selection paradigm. PhD Diss., University of Ljubljana, Faculty of Electr. Eng. and Computer Science, 1993.
[29] Leonardis, A., Gupta, A., Bajcsy, R.: Segmentation as the search for the best description of the image in terms of primitives. In: Proc. Third Int. Conf. on Computer Vision, pp. 121–125, Osaka, Japan, 1990. London: IEEE.
[30] Leonardis, A., Solina, F., Macerl, A.: A direct recovery of superquadric models in range images using recover-and-select paradigm. In: Proc. Third European Conf. on Computer Vision, pp. 309–318, Stockholm, Sweden, 1994.
[31] Loria, G.: Spezielle algebraische und transzendente ebene Kurven. Leipzig, Berlin: B. G. Teubner 1910.
[32] Maver, J., Bajcsy, R.: Occlusions as a guide for planning the next view. IEEE Trans. Pattern Anal. Machine Intell. *15*, 417–433 (1993).
[33] Marr, D., Nishihara, K.: Objects, parts, and categories. Proc. R. Soc. London Ser. *B 200*, 269–294 (1978).
[34] Metaxas, D., Dickinson, S.: Integration of quantitative and qualitative techniques for deformable model fitting from orthographic, perspective, and stereo projections. In: Proceedings of International Conference on Computer Vision, pp. 641–649. London: IEEE 1993.
[35] Muraki, S.: Volumetric shape description of range data using "Blobby model". Comput. Graphics *25*, 227–235 (1991).
[36] Nishihara, H. K.: Intensity, Visible-Surface, and Volumetric Representations. Art. Intell. *17*, 265–284 (1981).
[37] Pentland, A. P.: Perceptual organization and the representation of natural form. Art. Intell. *28*, 293–331 (1986).
[38] Pentland, A. P.: Recognition by parts. In: Proceedings of the First International Conference on Computer Vision, pp. 612–620. London: IEEE 1987.
[39] Pentland, A. P.: Automatic extraction of deformable part models. Int. J. Comput. Vision *4*, 107–126 (1990).
[40] Pentland, A. P., Sclaroff, S.: Closed-form solutions for physically based shape modeling and recognition. IEEE Trans. Pattern Anal. Machine Intell. *13*, 715–729 (1991).
[41] Ponce, J., Chelberg, D., Mann, W. B.: Invariant properties of straight homogeneous generalized cylinders and their contours. IEEE Trans. Pattern Anal. Machine Intell. *11*, 951–966 (1989).
[42] Raja, N. S., Jain, A. K.: Recognizing geons from superquadrics fitted to range data. Image Vision Computing *10*, 179–190 (1992).
[43] Schudy, R. B., Ballard, D. H.: Towards an anatomical model of hearth motion as seen in 4-D cardiac ultrasound data. Proceedings 6th Conference on Computer Applications in Radiology and Computer-Aided Analysis of Radiological Images, 1979.
[44] Solina, F.: Shape recovery and segmentation with deformable part models. PhD. Diss., University of Pennsylvania, Department of Computer and Information Science, 1987.
[45] Solina, F., Bajcsy, R.: Range image interpretation of mail pieces with superquadrics. In: Proceedings AAAI-87, Vol. 2, pp. 733–737. Seattle: WA, 1987.
[46] Solina, F., Bajcsy, R.: Recovery of mail piece shape from range images using 3-D deformable models. Int. J. Res. Eng. (Inaugural Issue) 125–131 (1989).
[47] Solina, F., Bajcsy, R.: Recovery of parametric models from range images: The case for superquadrics with global deformations. IEEE Trans. Pattern Anal. Machine Intell. *12*, 131–147 (1990).
[48] Stevens, K. A.: The visual interpretations of surface contours. Art. Intell. *17*, 47–73 (1981).
[49] Terzopoulos, D., Witkin, A., Kass, M.: Symmetry-seeking models and 3D object recognition, Int. J. Comput. Vision *1*, 211–221 (1987).
[50] Terzopoulos, D., Metaxas, D.: Dynamic 3D models with local and global deformations: Deformable superquadrics. IEEE Trans. Pattern Anal. Machine Intell. *13*, 703–714 (1991).
[51] Ulupinar, F., Nevatia, R.: Perception of 3-D surfaces from 2-D contours. IEEE Trans. Pattern Anal. Machine Intell. *15*, 3–18 (1993).
[52] Vidmar, A., Solina, F.: Recovery of superquadric models from occluding contours. In: Theoretical

foundations of computer vision (Klette, R., Kropatsch, W., eds.), pp. 227–240. Berlin: Akademie Verlag, 1992 (Mathematical Research, Vol. 69).
[53] Weiss, I.: 3-D shape representation by contours. Comput. Vision Graphics Image Proc. *41*, 80–100 (1988).
[54] Whaite, P., Ferrie, F. P.: From uncertainty to visual exploration. IEEE Trans. Pattern Anal. Machine Intell. *13*, 1038–1049, October 1991.
[55] Whaite, P., Ferrie, F. P.: Active exploration: Knowing when we're wrong. In: Proceedings of the ICCV, pp. 41–48, Berlin, Germany, 1993. IEEE CS.
[56] Zerrough, M., Nevatia, R.: Segmentation and recovery of SHGCs from a real intensity image. In: Proceedings Third European Conference on Computer Vision, Vol. 1, pp. 319–330. Stockholm, Sweden, 1994.
[57] Yokoya, N., Kaneta, M., Yamamoto, K.: Recovery of superquadric primitives from a range image using simulated annealing. In: Proc. 11th Int. Conf. on Pattern Recognition, Vol. A, pp. 168-172. IAPR, 1992. IEEE CS.

Dr. F. Solina
Computer Vision Laboratory
Faculty of Electrical Engineering
and Computer Science
University of Ljubljana
Tržaška c. 25
SI-61001 Ljubljana
Slovenia
e-mail: Franc.Solina@Ger.uni-G.si